コンピュータビジョンと
イメージメディア 103－13
（１９９７．１．２４）

表情認識とその程度推定

木村　聡　　谷内田　正彦

イメージ情報科学研究所　　大阪大学基礎工学部

本研究では人間の感情を表すいくつかの表情について、その認識、分類のみならず、無表情から最大まで表情の程度をおおまかに推定することを目的とした。本手法は、顔全体を１つのパターンとして捉え、無表情からの表情変化を抽出することにより表情認識が行えるとの考えに基づいている。表情程度のように微妙な情報を抽出するには、顔の造作の違いによる個人性を除去する必要がある。顔のエッジ画像からポテンシャルネットと呼ぶ物理モデルを用いて、表情変化は変形したネットとして表わされる。次に、ネットの変形をKL展開を用いて低次元の固有空間で表わすと、原点を平常顔としてその表情空間上の位置から表情程度を推定することが出来る。本研究では、”笑い”,”驚き”,”怒り”の３つの表情についてモデルを作成し、入力顔画像に対して評価を行なった。

Facial Expression Recognition and its Degree Estimation

Satoshi KIMURA* and Masahiko YACHIDA**

*Laboratories of Image Information Science and Technology

Senri LC 11F, 1-4-2, Shinsenri-Higashimachi, Toyonaka, Osaka, Japan 565

e-mail : kimura@image-lab.or.jp

**Faculty of Engineering Science, Osaka University

1-3 Machikaneyama, Toyonaka, Osaka, Japan 560

*e-mail : yachida@sys.es.osaka-u.ac.*jp

The purpose of this study is not only to recognize some kind of facial expressions which is associated with human emotion but also to estimate its degree. Our method is based on the idea that facial expression recognition can be achieved by extracting a variation from expressionless face with considering face area as a whole pattern. For the purpose of extracting subtle changes in the face such as the degree of expressions, it is necessary to eliminate the individuality appearing in the facial image. Using a physical model, we call Potential Net, a variation of facial expression is represented as the deformed Net from a facial edge image. Then, applying K-L expansion, the change of facial expression represented as the Net deformation is mapped into low dimensional eigen space, and estimation is achieved by projecting input images on to the Emotion Space. We have constructed three kind of expression models: happiness, anger, surprise, and experimental results are evaluated.

1.まえがき

視覚情報は人間同士のコミュニケーションにおいて中心的な役割を果たしている。ヒューマン情報処理において、顔は体の部位の中で最も重要で多くの情報を含んでいると言える。したがって、これまで顔画像認識に関する研究、特に表情認識に関する研究には多くの関心が払われており、人間とコンピュータとのインタラクティブな対話実現には不可欠な技術であると考えられる。

表情認識に関する研究の一般的な手法としては顔の特徴点、つまり目,眉,口などの器官の幾何学的位置や形状の変化を捕らえ、分析することにより認識を行う方法が多い。[1][2][3] しかしながら、それら特徴点の正確な位置や形状を異なる人物や、照明条件の異なる実画像から抽出することは非常に難しい。また、最近では動き情報に着目しオプティカルフローを用いる方法もある。[4][5] この方法では、連続画像から顔面筋や器官の動きをオプティカルフローとして抽出し、その時系列情報を分析することにより表情認識を行う。しかしながら、この方法の成否は抽出したオプティカルフローの信頼性に大きく左右され、顔のように複雑な剛体の動きからフローを正確に抽出することは難しい。

これに対して、我々の提案する方法は顔全体を1つのパターンとして捉える方法であり、特徴点の抽出や追跡を行わない。顔のエッジ画像全体をポテンシャル場と見なし、ポテンシャルネットと呼ぶ物理モデルを用いて顔画像から顔パターン全体をサンプリングする。[6][7] 松野ら[7]の研究においては、このネットモデルを用いて4つの表情モデル（幸福、驚き、怒り、悲しみ）を作成した。これらは、明白に表情を表わしている複数の人物の顔画像からネットを作成し、これを平均したものを各表情のモデルネットとしている。しかしながら、このような方法では明白な表情は認識出来るが、表情程度のように微妙な変化情報は、顔器官など顔の造作の個人性に埋もれてしまい抽出することは難しい。個人性による顔器官の幾何学的位置のバラツキは、表情変化による顔器官の位置変化よりも大きいと考えられる。本論文では、より微妙な表情の認識を目的とし、表情の程度を平常顔からの変位として推定出来るとの考えに基づく本手法を提案する。無表情からの表情変化のみによる特徴点の動きを捉えることにより、推定結果は個人の顔器官の位置に左右されない。したがって、ポテンシャルネットの変形で表された動きのパターンから表情を分類し、動きの変位から表情の程度を推定することができる。ポテンシャルネットの変形は表情の変化を表しており、これをKarhunen-Loeve展開を用いて主成分のパターン空間上に表し分析を行う。ネットの変形はノードの2次元変位ベクトルの集合で表され、そのままでは高次元であり扱いにくい。そこで、KL展開を用いて次元圧縮を行い、低次元の主成分空間上に表情程度のパターンを表した。これを感情空間と呼ぶ。入力画像に対しては同様にネットの処理を行い、感情空間上に写像を行うことにより表情の分類と程度推定を行う。

2．表情のサンプリング

2.1 顔画像の正規化

処理に用いる顔画像は、後で用いるネットのサイズに合わせるために、切り出し位置、大きさの正規化を行なう。切り出しの目印になる点としては、表情が変化しても顔面上でほぼその位置が一定している目、口の中心を基準とする方法を用いた。図1に示すように目、口の中心を基準として顔のほぼ中心を一定の大きさに照合領域を切り出す処理を行なった。以下にその手順を示す。

1．濃淡画像である顔正面像から左右の目と口の中心を手動により検出し、それぞれEr,El,Mとする。
2．線分$\overline{ErEl}$の中点をOとし、Mと結び線分$\overline{OM}$とする。
3．線分$\overline{ErEl}$と線分$\overline{OM}$がある一定の長さになるようアフィン変換を施す。
4．O点を基準に顔器官が領域からはみ出さないよう一定のサイズ(90×97)に切り出しを行なう。

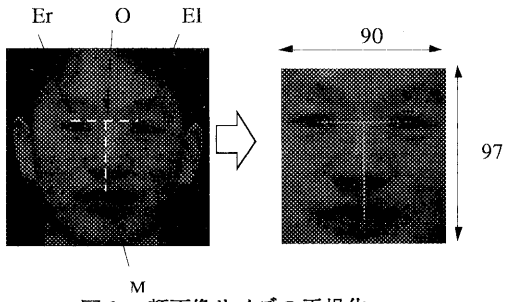


図1　顔画像サイズの正規化

各パラメータの値は、顔領域以外の背景や髪が含まれないように、出来るだけ大きな領域が取れる

よう経験的に設定した。上記の３の処理を行なうことにより、切り出し画像中での目、口の中心位置が一定になり、個人による顔器官位置のずれを吸収することが期待できる。

2.2 ポテンシャルネットモデル

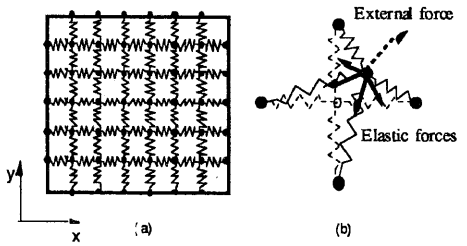


図２　ポテンシャルネットモデル

ポテンシャルネットは図２に示すように２次元グリッド状に各ノードをバネにより結合した物理モデルである。最外郭のノードは固定されており、他のノードは外力および結合されたバネの弾性力により移動する。従って、あるノードの位置変化はバネを伝わってネット全体の形状に影響を与え、影響の度合はバネ定数を変化させることによりコントロールすることが出来る。

表情の抽出にポテンシャルネットモデルを用いるのは、主に次の２つの理由からである。まず１つ目は、顔のエッジ画像はそのままではノイズを含んでおり、ポテンシャルネットを用いることにより表情変化には関係のない微細なノイズをある程度除去することが期待できる。２つ目に、顔の特徴点の動きは顔全体のパターンとしてポテンシャルネットにより抽出されるので、顔の器官を追跡する必要がない。

正規化を行なった濃淡画像からソーベルオペレータを用いて微分画像を作成し、ガウシアンフィルタ（σ=5）を用いてぼけ変換を行なう。この処理により図３に示すような顔器官やシワなどのエッジ付近に滑らかに勾配を持つ画像を作る。この顔のエッジ画像をポテンシャル場とみなす。これにポテンシャルネットを置き、場からの力を受けて変形させる。変形したネットはそのポテンシャル場全体を構造的にサンプリングしており、顔全体を１つのパターンとして表わしている。

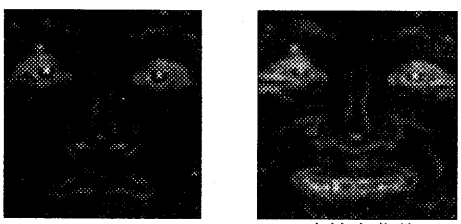


図３　ポテンシャル場としての顔エッジ画像

2.3 ポテンシャルネットの変形

ネットのノード$N(i,j)$の動きは下記の方程式により支配される。

$$m\frac{d^2n_{i,j}}{dt}+\gamma\frac{dn_{i,j}}{dt}+F_{spring}=F_{ext} \qquad (1)$$

ここで、$N(i,j)$は２次元の座標、mはノードの質量、γは減衰係数、F_{spring}は内力（弾性力）、F_{ext}は外力（画像からの力）である。ここでは、問題を簡単化するため運動方程式（１）を静的問題として捉え、その平衡状態を考える

$$F_{spring}=F_{ext} \qquad (2)$$

式(2)は弾性力と画像からの力の平衡状態を表している。

ノード$N(i,j)$に働く弾性力は下記のように表される。

$$F_{spring}=k\sum_{a}^{4}\left(\left|l_{i,j}^{a}\right|-l_0\right)\frac{l_{i,j}^{a}}{\left|l_{i,j}^{a}\right|} \qquad (3)$$

ここで、lはバネの長さ、kは弾性係数、l_0は隣接したノード間の距離の初期値である。

ノード$N(i,j)$における外力は下記のように表される。

$$F_{ext}=\alpha\left(\nabla\left(G_\sigma * I\left(x_{i,j},y_{i,j}\right)\right)\right) \qquad (4)$$

ここで、αは外力を制御する係数、Iはノード$N(i,j)$におけるエッジ画像の濃度値、またGは幅σをもつ２次元ガウシアンフィルタによるコンボリューションを表す。

式(2)の安定状態にネットを収束させるため、各ノードを８方向に動かし、下記の合力が最小となる方向へ移動させる。

$$F_{total}=F_{ext}-F_{spring} \qquad (5)$$

上記の処理を繰り返すことにより、全ノードの合力F_{total}がそれぞれ閾値以下になったところで、ネットが平衡状態にあるとみなす。

2.4 表情変化の抽出

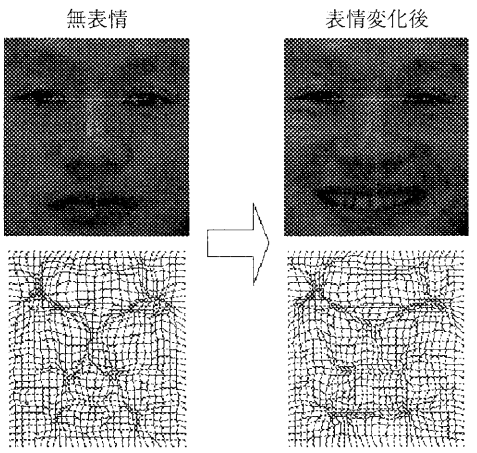


図4　基準ネットからの変形

本研究では感情を表すある表情について、無表情から最大まで表情の程度をおおまかに推定することを目標としている。表情程度のように微妙な変化を捕らえるにはモデル個人の基準となる表情である平常顔からの変化分を抽出し、分析することにより認識を行うというのが本手法の基本的な考え方である。無表情からの表情変化に起因する特徴点の動きを抽出することにより、顔器官の幾何学的配置の違いによる個人性に影響されることなく認識を行うことが出来る。したがって、ネットの変形パターンから表情の分類を行い、変形の程度から表情程度を推定することが出来る。無表情の顔のポテンシャルネットを基準として用い、基準ネットからのネットの変形として表情の変化を抽出することが出来る。図4に示すように、まず無表情の顔画像から基準ネットを作成する。これを表情の現われた顔画像に被せて変形させてネットを作成する。そして式(6)に示すように、基準ネットからの変位をネットの各ノードの2次元変位ベクトルとして抽出する。

$$V_{i,j} = n^{e}_{i,j} - n^{0}_{i,j} \qquad (6)$$

$V_{i,j}$：変位ベクトル
$n^{0}_{i,j}$：グリッドの初期位置
$n^{e}_{i,j}$：グリッドの変位位置

以上のように求めたネットの全グリッドの変位ベクトルが表情の変化を表している。

3．表情程度の推定

3.1 表情程度モデル

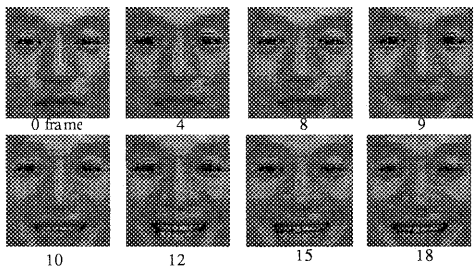


図5　表情程度モデル

心理学における研究では人間には万国共通で、感情を表す6種類の基本的表情があることが報告されており、多くの表情認識に関する研究もこの6種類の表情の認識を目的としている。本研究ではこの内の3種類の表情の”笑い”,”怒り”,”驚き”について表情の分類のみならず、各表情の程度を推定することを課題としている。従って、まずこれら表情の程度を表すモデル顔画像を用意する必要がある。そこで、本研究では各表情について平常顔から徐々にその表情に変化してゆく連続画像の各フレームを表情程度モデルとして用いた。図3にモデル画像の例を示す。

3.2 ポテンシャルネットの　KL展開

我々は、ポテンシャルネットの変形を顔全体の特徴として抽出している。しかしながら、ネットの変形を表すグリッドの変位ベクトルは、本研究で用いているネットでは29×31個となり、そのままでは扱いにくい。そこで、パターン認識における統計的特徴抽出法として知られているＫＬ展開を用いることにより、情報圧縮を行い低次元のベクトル空間で表情の特徴を表現することができる。以下にその処理を述べる。

1．3種類の表情程度モデル画像から作成した合計n個の表情モデルネットM_dの平均μを求める。

2．共分散行列Cをもとめる。

$$C = \frac{1}{n}\sum_{d=1}^{n}(M_d - \mu)(M_d - \mu)^T = AA^T \qquad (7)$$

3．共分散行列Cの固有値λ_fと固有ベクトル

U_fを求める。

$$CU_f = \lambda_f U_f \qquad (8)$$

$$U_f^T U_f = 1 \qquad (9)$$

ここで固有ベクトルU_fは下記のように求められる。

$$A^T A W_f = \Phi_f W_f \qquad (10)$$

$$AA^T A W_f = \Phi_f A W_f \qquad (11)$$

$$U_f = A W_f \qquad (12)$$

ここで行列$A'A$の固有値をΦ_f、固有ベクトルをW_fとする。

4．固有値の大きい順にG個の固有ベクトルを照合用として用いる。

5．G次元ベクトル$R_{df} = (r_{d1}, r_{d2}, ..., r_{dG})^t$が表情程度モデル$M_d$に対する特徴ベクトルとなる。

$$r_{df} = U_f^T (M_d - \mu) \quad f = 1, ..., G \qquad (13)$$

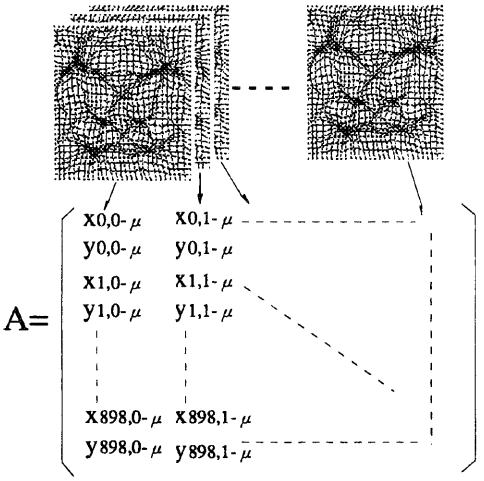
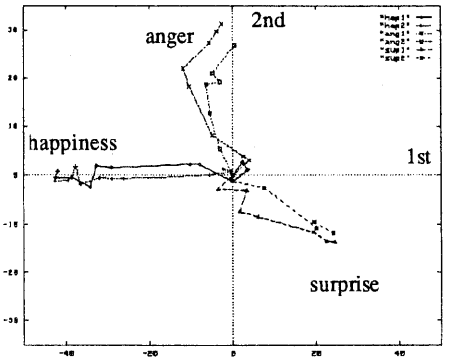


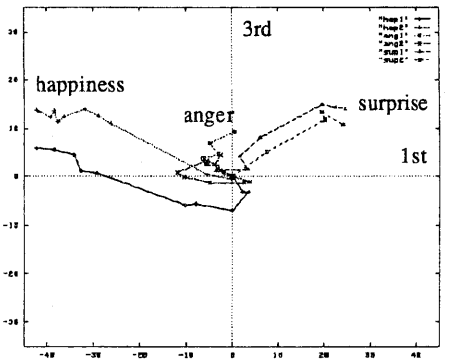
図6　ポテンシャルネットのＫＬ展開

3.3 感情空間

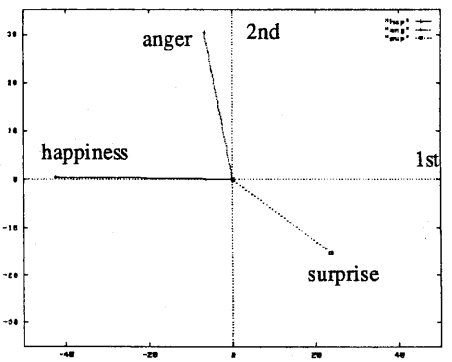
”笑い”,”怒り”,”驚き”の3種類の表情について、各表情程度モデル画像からネットを作成し、平常顔ネットからの変化分をグリッドの変位ベクトルとして取り出す。これを成分としてＫＬ展開を用いて主成分分析を行う。図7は上位の2成分を軸にとった固有空間であるが、平常顔の位置を原点として3表情の程度モデルの軌跡が描かれ、この空間上の位置から表情の分類と程度の推定を行なうことがきる。この固有空間を我々は感情空間と呼ぶ。ここでは認識処理を簡単にするため、図7(c)に示すように第一,第二主成分で表されたモデルの軌跡を最小二乗法により線分で表している。つまり、この線分の方向が表情の種類を示し、原点からの距離が程度を表している。入力画像はモデルの線分に対して原点からの方向によ



(a) 1st-2nd 主成分空間



(b) 1st-3rd 主成分空間



(c) 表情程度近似モデル

図7　感情空間

り表情を分類し、距離により表情程度を推定することが出来る。

4．実験

4.1 モデル空間の作成

表情程度を表すモデルとして平常顔から徐々に表情が変化してゆく顔連続画像を用いた。3種類の表情についてそれぞれ2つずつの連続顔画像を用意し、これを表情程度モデルとして用いた。各連続画像はそれぞれ平常顔から最大の表情に到達するまでの時間が異なっており、モデルに用いた画像での画像のフレーム数は下記のようになった。

笑い：10 frame, 13 frame

怒り：7 frame, 9 frame

驚き：8 frame, 8 frame

モデルの軌跡を最小二乗法により線分で近似しこれを各表情の照合パターンとして用いる。ここでは上位2次元までの主成分を照合に用いたが、この時の累積寄与率は表1の通り32%である。

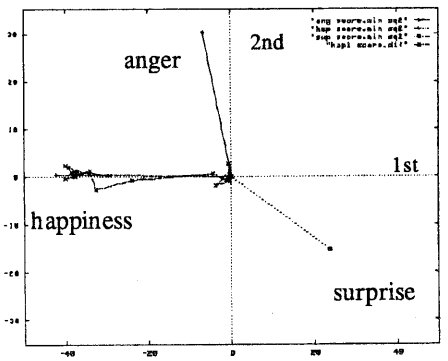
基底	1st	2nd	3rd	4th	.....
累積寄与率(%)	21	32	37	41	.....

表1　累積寄与率

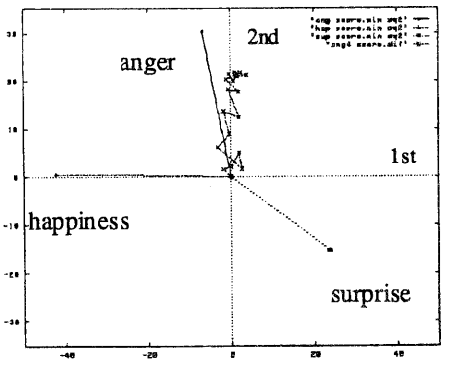
4.2 入力画像の認識

まず最初の実験では、モデルと同様に平常顔からある表情へと徐々に変化してゆく。3種類の各表情ごとに10個の、長さがそれぞれ20frameの連続画像を用意した。ここで、これらの画像はモデルと同一人物である。図8に入力画像をモデル空間に投影した例を示す。入力画像の20frame中に実際に表情が変化している部分は、モデルと同様に10frame程度であり、始めと終わりにはあまり表情変化のない部分があることが分かる。入力画像の軌跡は、いずれもモデルの線分に沿って原点から外に向かって延びてゆく。実画像と同様に始めと終わりにあまり変化のない部分があることが分かる。この様に、未知入力画像に対して、そのモデル空間上の位置から表情の分類と程度の推定を行うことが出来る。図8から分かるように表情の急激な変化は、ほんの数フレームで起こっている。実験に用いた合計30個の入力画像の内、表情分類については100%の認識結果を得た。表情程度の推定については、モデル空間上では大,中,小の3段階ほどの程度分類が可能であると思われる。

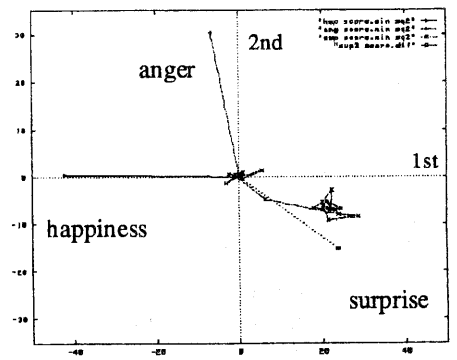
次に未知の人物の入力画像に対する実験結果例を図9に示す。図9に示した顔画像はモデルの人物とは異なっており、3種類の表情を表出している。これらを感情空間上に写影し認識を行った。グラフに示されているように各表情はほぼ適当な位置にあるが、いくつかのケースで不適当な位置にくる場合があった。



(a) 幸福



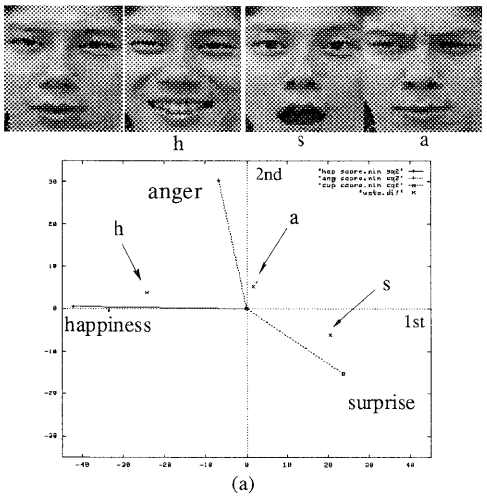
(b) 怒り



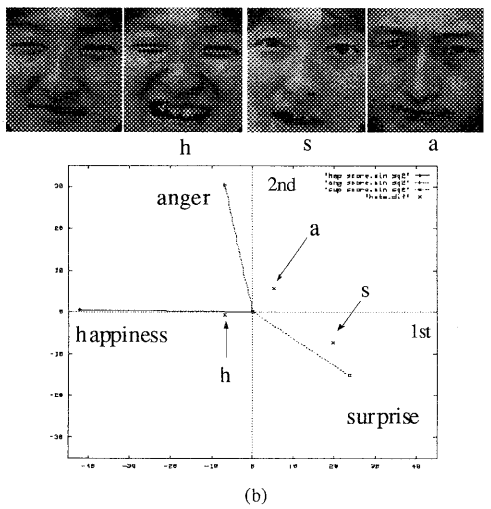
(c) 驚き

図8　入力画像の例

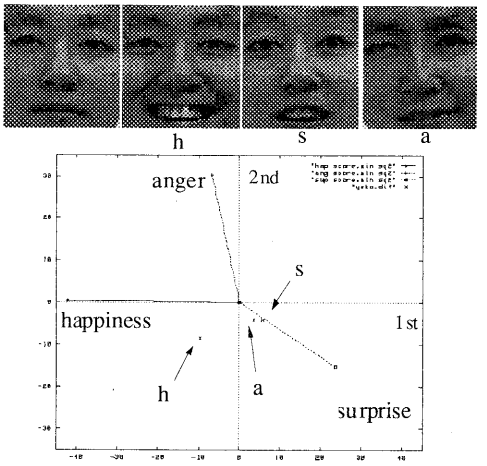
未知の人物に対する実験結果が思わしくないのは、いくつかの理由が考えられる。１つの理由としては、モデル空間を一人の人物の画像から作成したため、各表情を表出する時の顔器官の動作の方法に個人性があり、モデルの表情パターンに適合しなかったと考えられる。つまり、ある１つの表情に対しても、各個人により様々な表出の方法があると思われる。また他の理由として、カメラの前で意図的にある表情を表出する場合、本実験に用いたような普通の人物は、俳優のように上手に表情を作ることは難しい。従って、どうしても不自然な表情になりがちであり、少なからず認識結果に影響を与えると考えられる。



(a)



(b)



(c)

where h:幸福, s:驚き, a:怒り

図９　未知の人物の入力例

４．まとめ

本論文では、顔画像からの表情認識とその表情程度推定を行う新たな手法を提案した。本手法の基本的な考え方は、ある表情の程度は平常顔からの変化分として抽出することが出来るということである。表情変化の抽出にはエッジ画像からポテンシャルネットを用いて顔全体を１つのパターンとしてサンプリングを行った。抽出したネットの変形をＫＬ展開を用いて低次元のベクトル空間で表現し表情空間を作成した。表情程度を表すモデルには平常顔から徐々にある表情に変化してゆく連続画像を用いた。入力画像を表情空間上に投影し、原点からの向きにより表情を分類し、距離により程度を推定することが出来る。本手法は比較的簡便な方法で表情程度のように微妙な情報を推定することが出来る。

今後の課題としては、未知の人物に対する認識結果を向上させるため、複数の人物の顔画像からモデルを作成する必要があると思われる。また、実験に使う顔画像は意図的でなく自然に表出された表情の画像を用いる必要があると考えており、TVや映画から顔画像を取り出すことも検討している。

参考文献

[1]M.Suwa, N.Sugie and K.Fujimura, "A preliminary note on pattern recognition of human emotional

expression", in Proc.IJCPR, pp.408-410, 1978

[2]D.Terzpoulos and K.Waters, "Analysis and Synthesis of Facial Image Using Physical and Anatomical Models", IEEE trans. Patt. Anal. Machine Intell.,vol.15, No.6, pp.569-579, 1993

[3]M.A.Shackleton and W.J.Welsh, "Classification of Facial Feature for Recognition", Proc.CVPR, pp.573-579, 1991

[4]Y.Yacoob and L.Davis, "Computing Spatio-temporal Representations of Human faces", in Proc.CVPR, pp.70-75, 1994

[5]I.A.Essa and A.Pentland, ``Facial Expression Recognition using a Dynamic Model Energy", in Proc. of 5th ICCV, pp.360-367, 1995

[6]S.Kimura, C.W.Lee and M.Yachida, "Extended Facial Expression Recognition Using Potential Net", In Proc. of ACCV'95, Ⅲ pp.728-732, 1995

[7]K.Matsuno, C.W.Lee, S.Kimura and S.Tsuji, "Automatic Recognition of Human Facial Expressions", in Proc. of 5th ICCV, pp.352-359, 1995

[8]松野,李,辻,"ポテンシャルネットとKL展開を用いた顔表情の認識", 信学会論文誌D Ⅱ,Nc8, pp1591-1600,1994年8月

[9]D.Terzopoulos and M.Casilescu,"Sampling and Reconstruction wit Adaptive Meshes", in Proc. CVPR-91, pp.70-75, 1991

[10]P.Ekman and W.V.Friesen, "The Facial Action Coding System" ,Consulting Psychologist Press, Inc., SanFrancisco, CA, 1978

[11]M.Turk and A.Pentland,"Face Recognition Using Eigenfaces", in Proc.CVPR, pp.586-591, 1991

[12]H.Wu, T.Yokoyama, D.Pramadihanto and M.Yachida, "Face and Facial Feature Extraction from Color Image", Proc. of 2nd Int. Workshop on Auto Face and Gesture Recognition. pp345-350 1996